

# LOFAR Data Services: Now & Next

---

R. F. Pizzo  
Head Science Data Centre Operations

18 September 2024

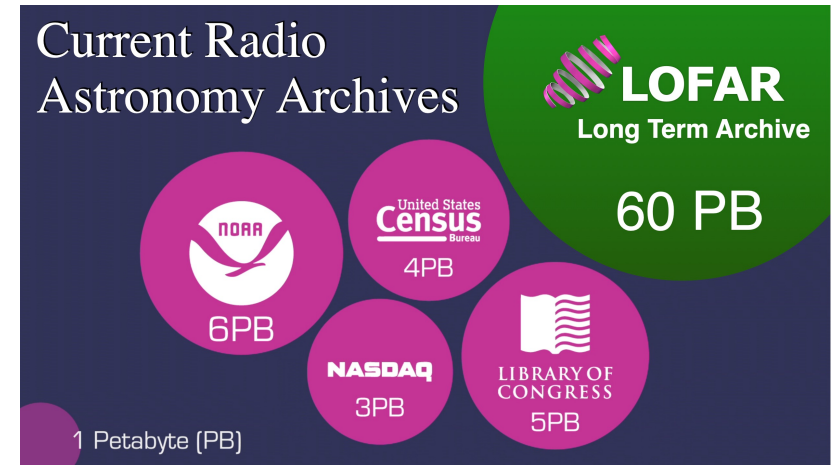


# Outline

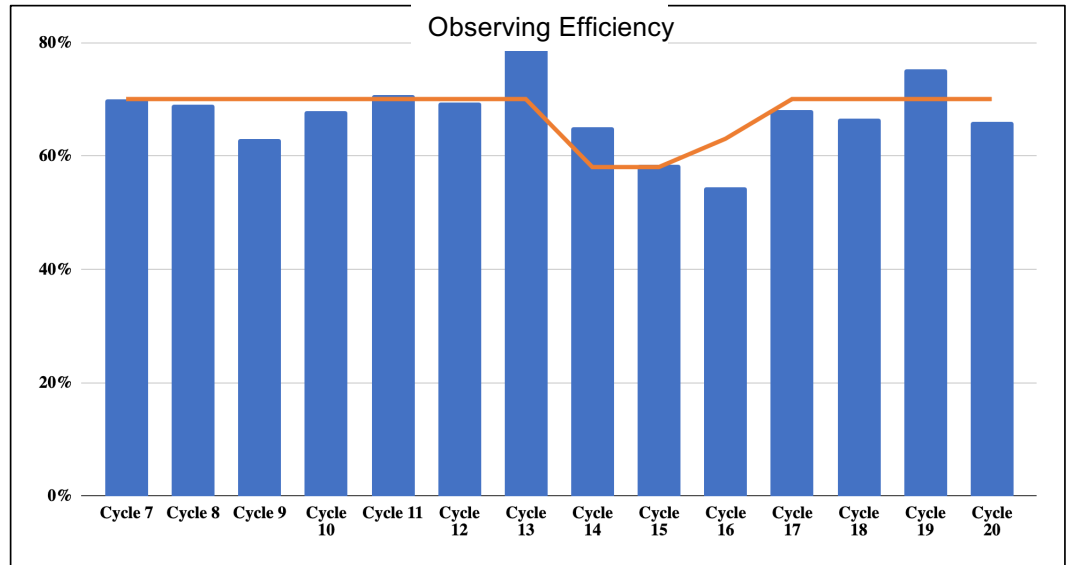
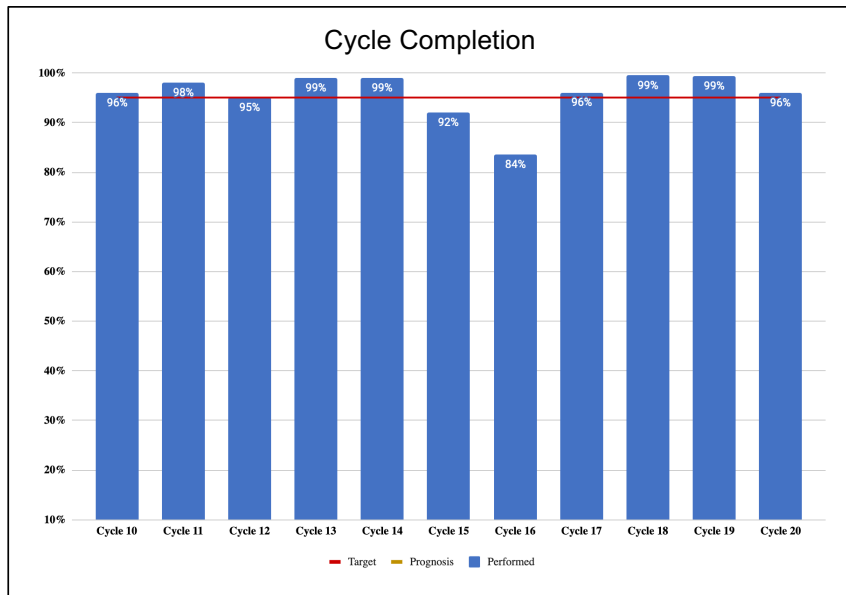
- Current status & successes
- Future ambition
- Challenges
- Resource challenges
- Data policies

# Achievements

- Completed **21 operational Cycles**
- **~65000** hours successfully observed – **>70% operational efficiency**
- Operating a massive array growing in size and capabilities
- **60 PB (!) in the LTA- Largest astronomical data collection to date.**
- **Started LTA operations**
- Supported an **ever-growing community**
- Brought the instrument closer to our users:
  - **LOFAR Schools** (400+ participants)
  - **60 Busy Weeks**
  - **Traineeships**

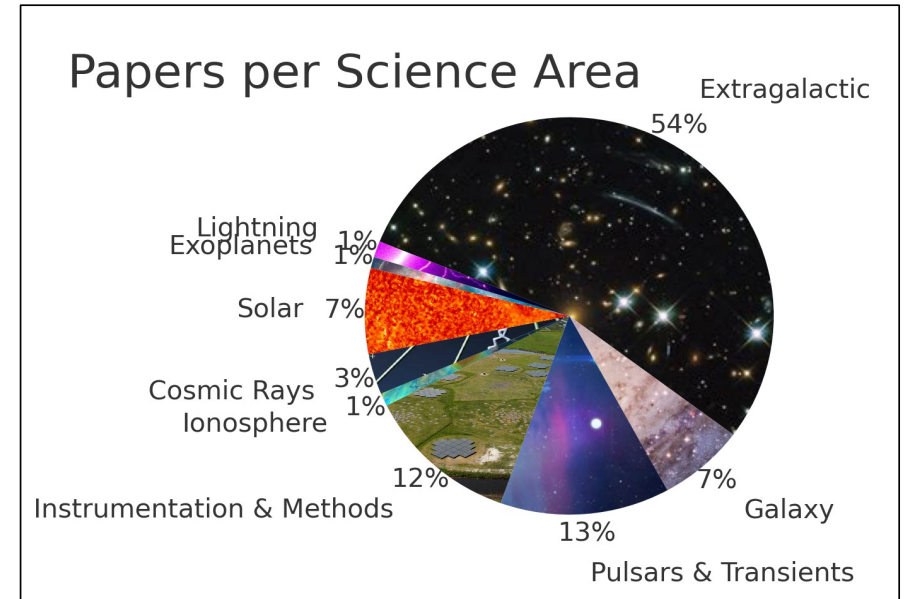
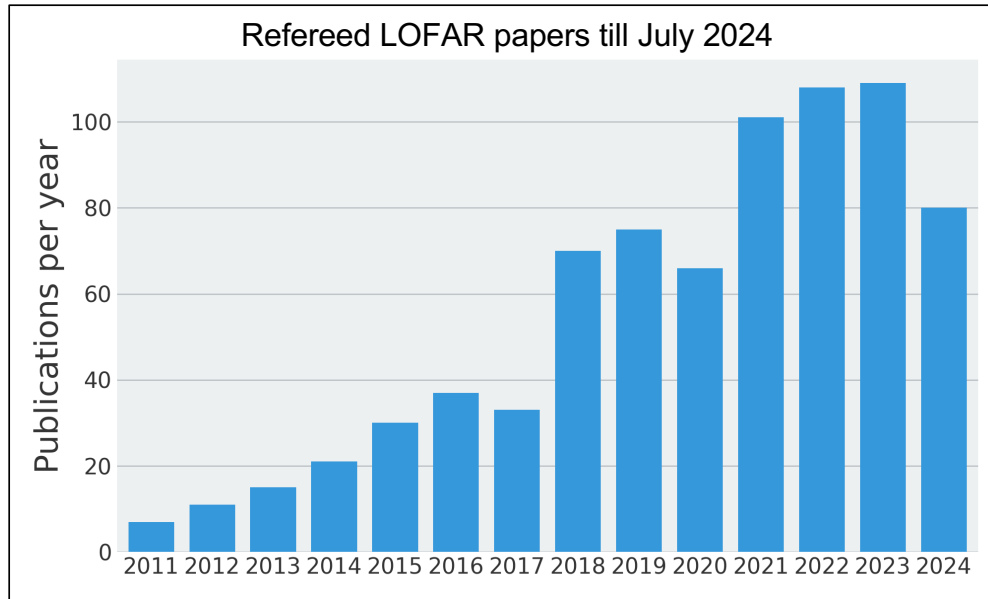


# Observatory Performance



- Target efficiency: 70%
- Target completion: 95%
- Cycle 20: 96% completed

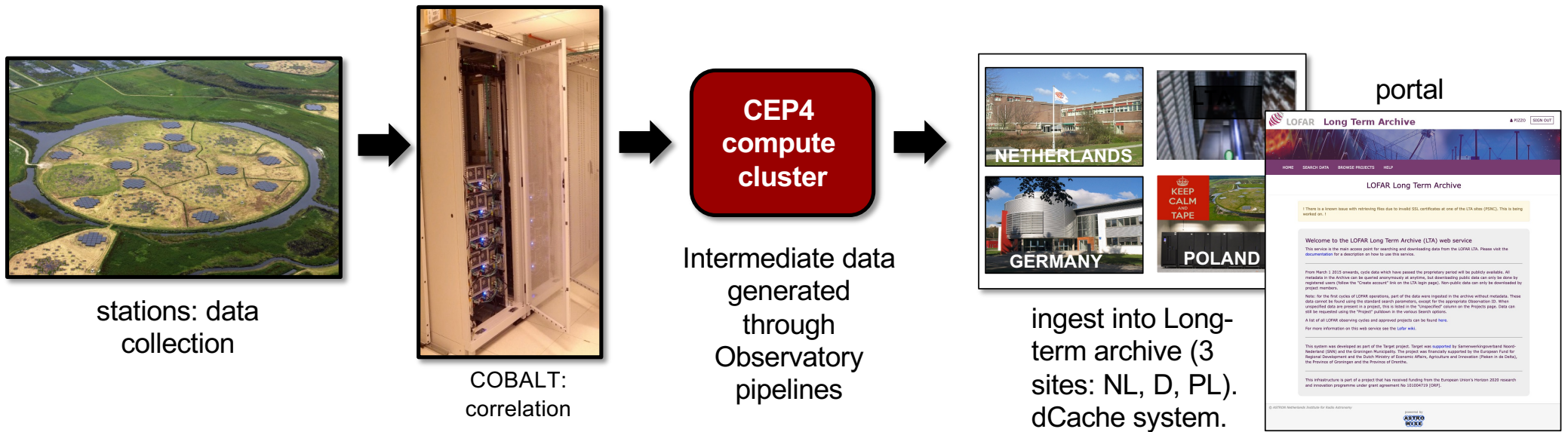
# Lofar Science Output



- > 750 refereed publications
- Publication rate: **2 papers per week** – top 10% of all astronomical facilities



# The LOFAR System: Data Flow...so far



## ➤ Transport, processing and storage of large amounts of data :

- Data flow from all antennas combined: 1.7 Tbyte/s
- To COBALT from station after beamforming: 28 Gbyte/s
- Correlator output to disk: between 2-10 Gbyte/s
- Data storage challenges: ~ 80 TB/h
- Data transfer to the archive: ~10 TB/h
- Archive now: ~ 60 PB in mixed state of reduction and science readiness

# CENTRALIZED PROCESSING OF LOFAR DATA...SO FAR

- Initial processing in Groningen
- Single copy of uncalibrated data on one of the three archive sites
- Advanced processing handled by PI's on external facilities (or at the LTA)
- Some advanced products released through data releases

**SURF REPOSITORY** HOME | USER GUIDE | FAQ Log in

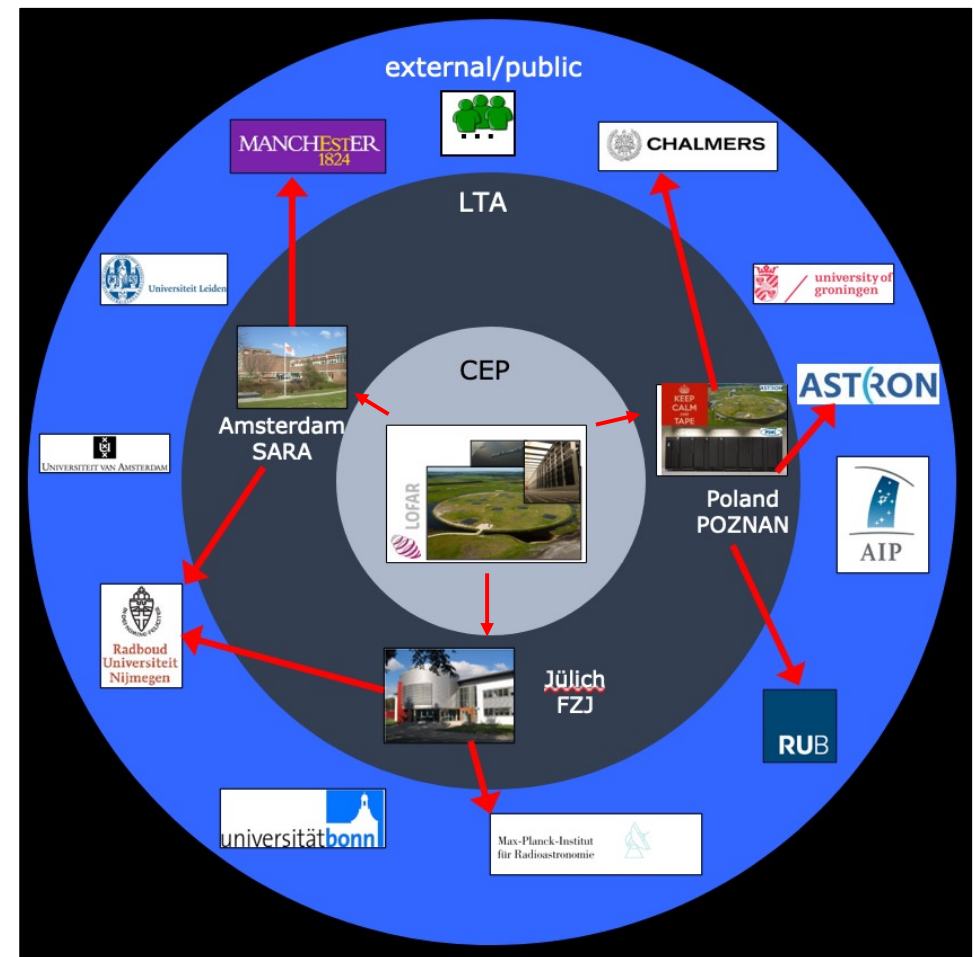
Export as ▾

**The LOFAR Two-metre Sky Survey (LoTSS) V. Second data release**

T. W. Shimwell, M. J. Hardcastle, C. Tasse, P. N. Best, H. J. A. Röttgering, W. L. Williams, A. Botteon, A. Drabant, A. Mechev, A. Shulevski, R. J. van Weeren, L. Bester, M. Brügger, G. Brunetti, J. R. Callingham, K. T. Chyży, J. E. Conway, T. J. Dijkema, K. Duncan, F. de Gasperin, C. L. Hale, M. Haverkorn, B. Hugo, N. Jackson, M. Mevius, G. K. Miley, L. K. Morabito, R. Morganti, A. Offringa, J. B. R. Oonk, D. Rafferty, J. Sabater, D. J. B. Smith, D. J. Schwarz, O. Smirnov, S. P. O'Sullivan, H. Vedantham, G. J. White, J. G. Albert, L. Alegre, B. Asabere, D. J. Bacon, A. Bonafede, E. Bonnassieux, M. Brienza, M. Bilicki, M. Bonato, G. Calistro Rivera, R. Cassano, R. Cochrane, J. H. Croston, V. Cuciti, D. Dallacasa, A. Danezi, R. J. Dettmar, G. Di Gennaro, H. W. Edler, T. A. Enßlin, K. L. Emig, T. M. O. Franzen, C. García-Vergara, Y. G. Grange, G. Gürkan, M. Hajduk, G. Heald, V. Heesen, D. N. Hoang, M. Hoeft, C. Horellou, M. Iacobelli, M. Jamroz, V. Jelc, R. Kondapally, P. Kukreti, M. Kunert-Bajraszewska, M. Magliocchetti, V. Mahatma, K. Malek, S. Mandal, F. Massaro, Z. Meyer-Zhao, B. Mingo, R. I. J. Mostert, D. G. Nair, S. J. Nakoneczny, B. Nikiel-Wroczyński, E. Orrú, U. Pajdosz-Smierciak, T. Pasini, I. Prandoni, H. E. van Piggelen, K. Rajpurohit, E. Retana-Montenegro, C. J. Riseley, A. Rowlinson, A. Saxena, C. Schrijvers, F. Szejten, T. M. Siewert, R. Timmerman, M. Vaccari, J. Vink, J. L. West, A. Wolowska, X. Zhang, J. Zheng

**1 January 2022**

In this data release from the ongoing LOW-Frequency ARray (LOFAR) Two-metre Sky Survey (LoTSS) we present 120-168 MHz images covering 27% of the northern sky. Our coverage is split into two regions centred at approximately 12h45m +44°30' and 1h00m +28°00' and spanning 4178 and 1457 square degrees respectively. The images were derived from 3,451 hrs (7.6 PB) of LOFAR High Band Antenna data which were corrected for the direction-independent instrumental properties as well as direction-dependent ionospheric distortions during extensive, but fully automated, data processing. A catalogue of 4,396,228 radio sources is derived from



# Challenges: Data Storage, Access, Distribution & Curation

## ➤ Storing data gets costly very quickly, especially online storage:

- disk- Online-Pb-Year ~ € 100,000
- tape - Nearline-Pb-Year ~ € 14,000
- one enters a regime where re-observing is cheaper

## ➤ Data access

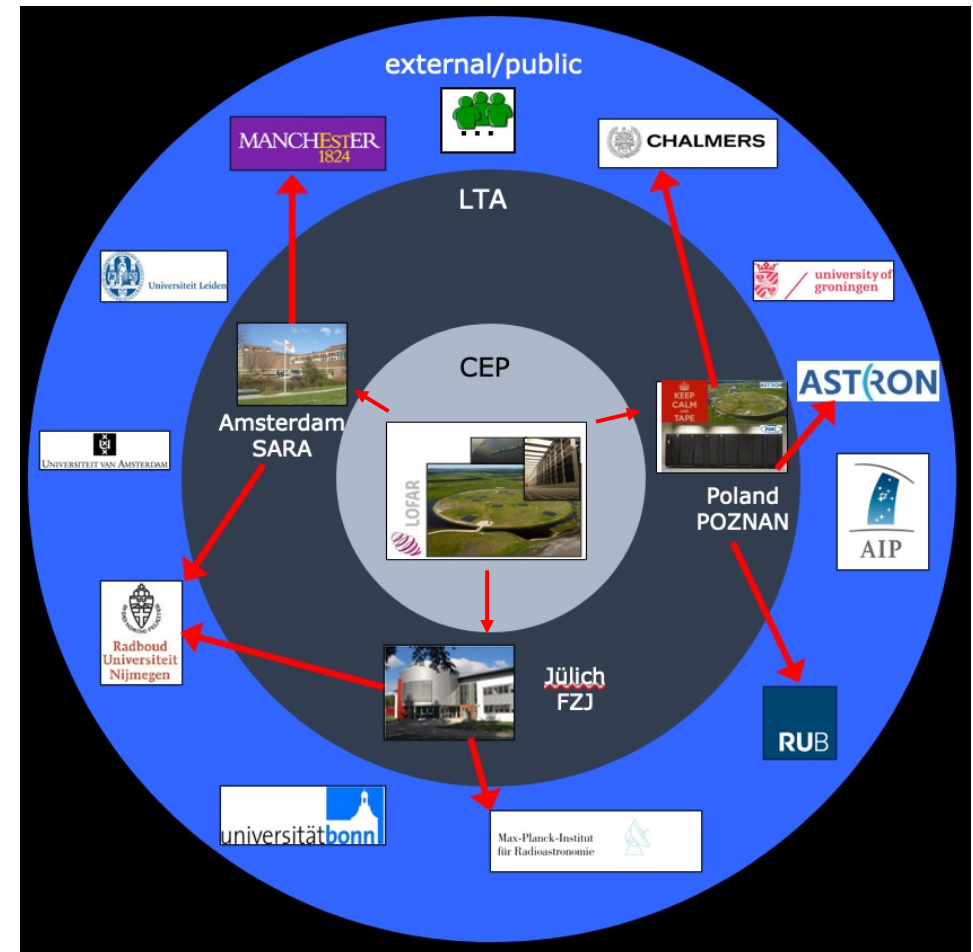
- Need good interfaces and functionalities that help users to mine the archive and find the data
- Intermediate data should be easily retrievable:
  - user data access limited by dCache overall capacity and bandwidth
  - LOFAR software distribution required

## ➤ Curation:

- Making available advanced products from users
  - Need 'user ingest' and a LOFAR data 'hub'

## ➤ Distribution

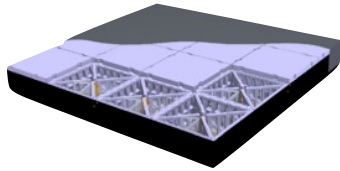
- Moving large amounts of data is impractical



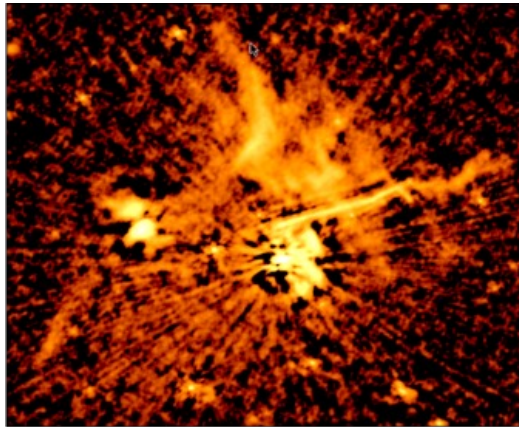


# COMING NEXT: LOFAR2.0

High-Band



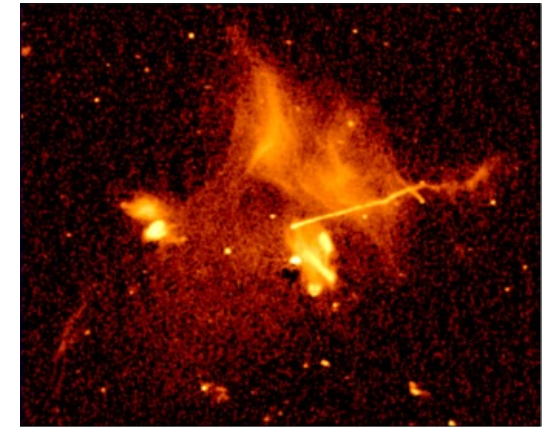
Scientifically limited



Breakthrough techniques



Rich in science



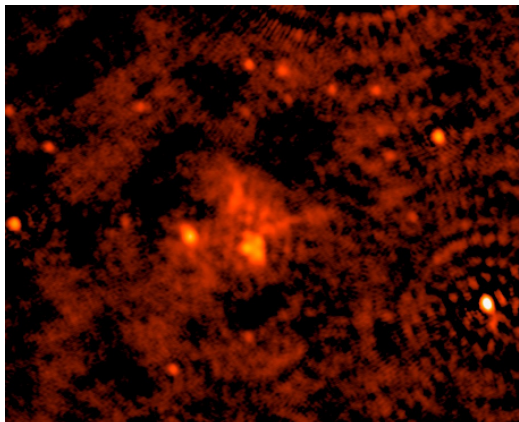
No ionospheric correction

Ionosphere well modeled

Low-Band



2x



LOFAR2.0

AUTOMATIC



The Goal

ASTRON

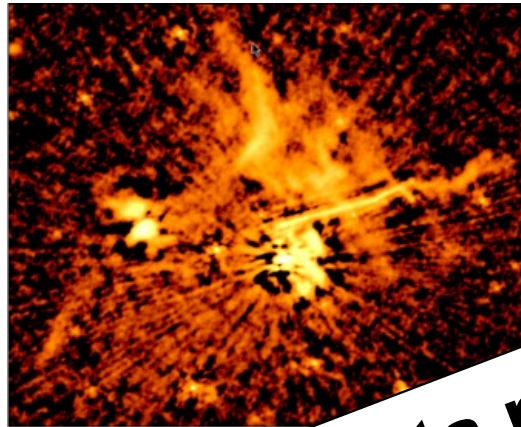
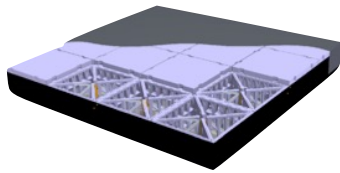
Netherlands Institute for Radio Astronomy

# COMING NEXT: LOFAR2.0

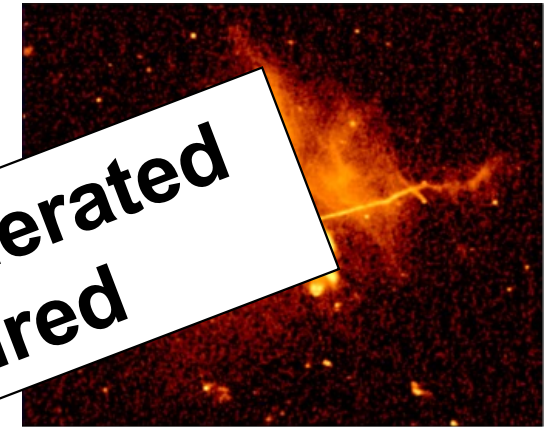
Scientifically limited

Rich in science

High-Band



Breakthrough techniques

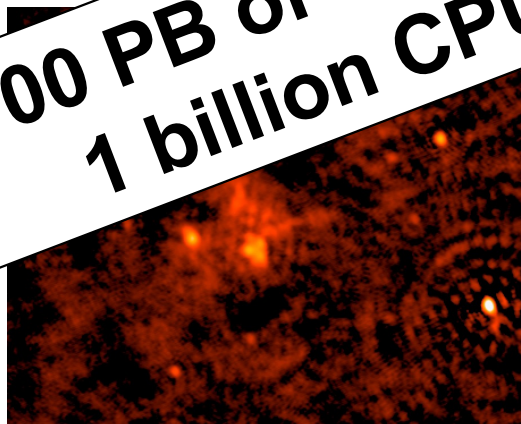


Low-Band



2x

No ionospheric



Ionosphere well modeled

**>100 PB of data products generated  
1 billion CPU hours required**

LOFAR2.0

AUTOMATIC

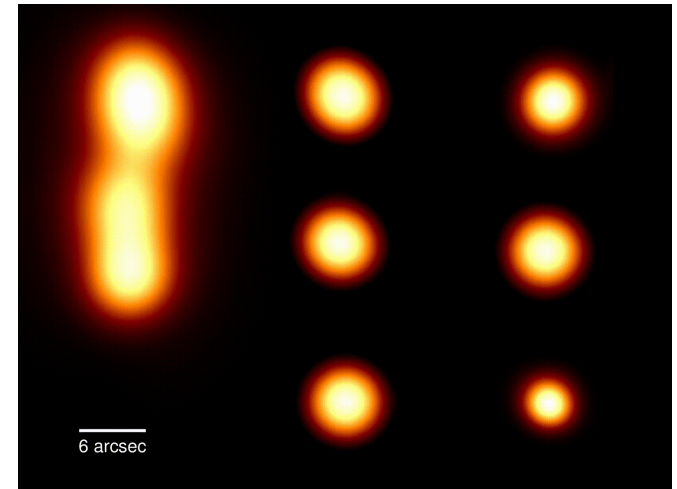
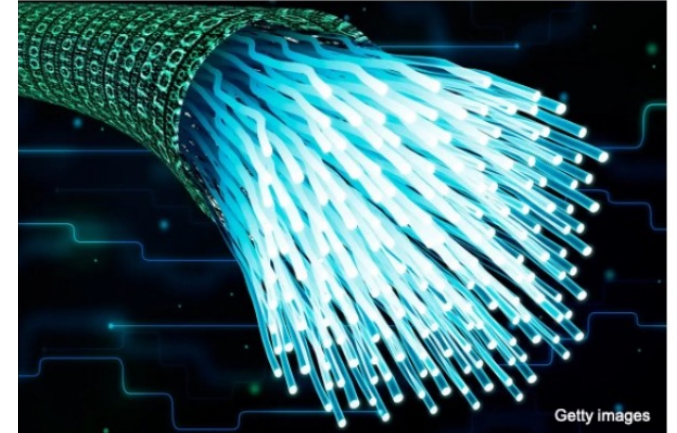


The Goal

**ASTRON**

Netherlands Institute for Radio Astronomy

- LENS – LOFAR Enhanced Network for Sharp Surveys
- Upgrade the network (10 → 100Gb/s) for full-FOV, full-res imaging
- Will require high-throughput data processing system deploying innovative algorithms capable of keeping up with the data streaming from the telescope
- Data products generated: **50 PB/year**
- Computing: **1 billion CPU hours**



# Tackling the Challenge: Generating Science-Ready Data

LOFAR Long Term Archive

HOME SEARCH DATA BROWSE PROJECTS HELP LOGIN

LC1\_027

Observation 1 to 100 (showing 100 of total 387) -

Averaging Pipeline (total 0) -

Calibration Pipeline (total 0) -

Imaging Pipeline (total 0) -

Long Baseline Pipeline (total 0) -

Pulsar Pipeline 1 to 100 (showing 100 of total 387) -

edit columns

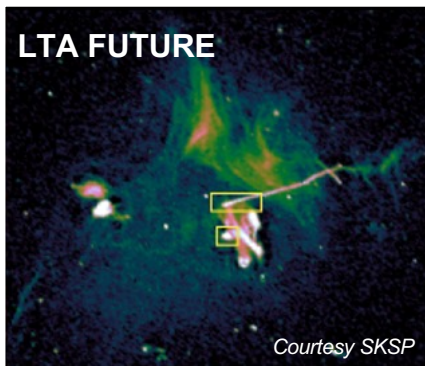
first previous 1 2 3 4 next last

#	Project	Release Date	Pipeline Name	Pipeline Version	SAS Id	Pulsar Selection	doSinglePulseAnalysis	Strategy Name	convertRawTo8bit [s]	subIntegrationLength	Source DataProduct	All Dataproducts	Quality	Pulsars
1	LC1_027	2015-05-15	J1544+4937/PULP	n/a	1027091	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	Good	0
2	LC1_027	2015-05-15	B1237+25/PULP	n/a	1027069	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	Good	0
3	LC1_027	2015-05-15	B1133+16/PULP	n/a	1027047	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	Good	0
4	LC1_027	2015-05-15	J1024-0719/PULP	n/a	1027025	Pulsars in observation specs, file or SAP	0	Pulsar Pipeline	0	-1.0	show	show	Moderate	0

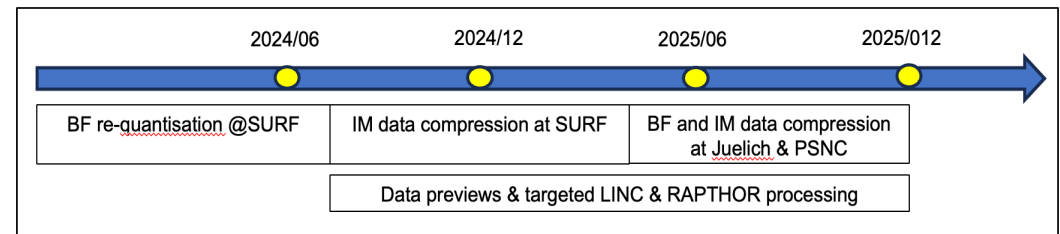
**data quality flag**

- New operational area compared to LOFAR1
- **LDV:**
  - **Reduce data volume** at the LTA to reduce operational costs
  - **Streamline data processing** operations at the LTA
  - Prepare ASTRON for **LOFAR2 Large Programs**
- **LDV Operations started early 2023**
  - Current focus: BF data processing (re-quantization) – IF data compression following this year. **Savings so far: >5PB** + introduction of data quality flag
- Further valorisation of portions of the archive through **data previews** (*marriage* with LOFAR2.0 commissioning)

LTA FUTURE



LTA FUTURE



LDV timeline



# Tackling the Challenges: Forthcoming Tools and Functionalities Developed by the SDC

Support for advanced products, become a 'hub' for access to LOFAR data

Proposal Management

Archiving & Curation

ATDB Dashboard Filter: Quality Validation Failures Discarded Finished Monitoring Diagram [Sign In](#)

Click to Filter 0

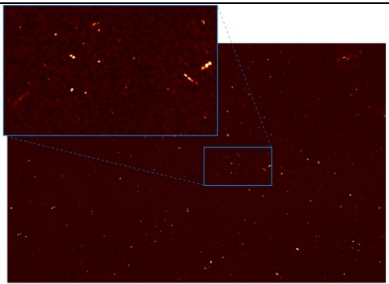
Clear Filter 0

Workflow: [selected] [archiving] [archived] [archive] [suspend]

First: Previous 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 Next Last

ID	Workflow	Priority	Status	Project	SAS_ID	Filter	CreationTime	Size	Actions
142970	▼	100	scrubbed (holding)	lta_005	666004	lta_005- rp-2b05	2023-11-01 09:32:05	20.2 MB	
142969	▼	100	scrubbed (holding)	lta_005	666004	lta_005- rp-2b05	2023-11-01 09:32:05	14.9 GB	
142968	▼	100	scrubbed (holding)	lta_005	666004	lta_005- rp-2b05	2023-11-01 09:32:05	17.8 GB	

For generation of science-ready data



Scientific Pipelines

LOFAR2.0 Digital Services

Managed Processing

Discovery & Access

Interactive Data Analysis  
**LATER**

Robust and reliable access to data, use of VO interfaces to publish data, FAIRness

User Pipeline Execution  
**LATER**

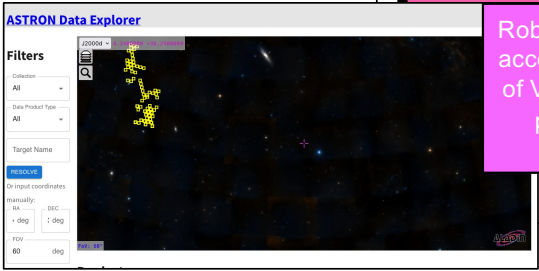


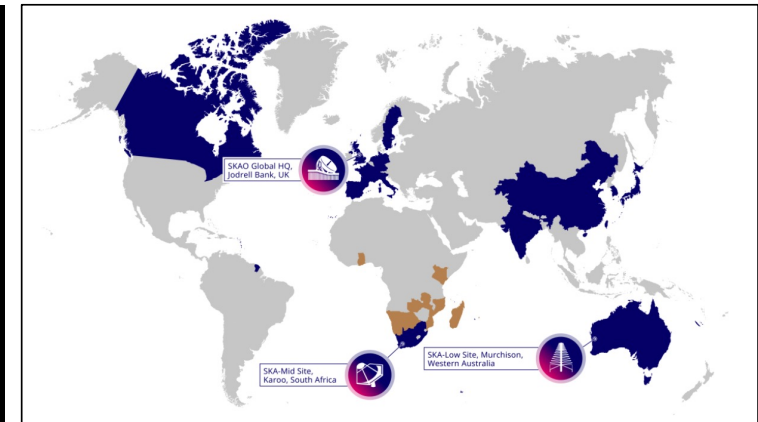
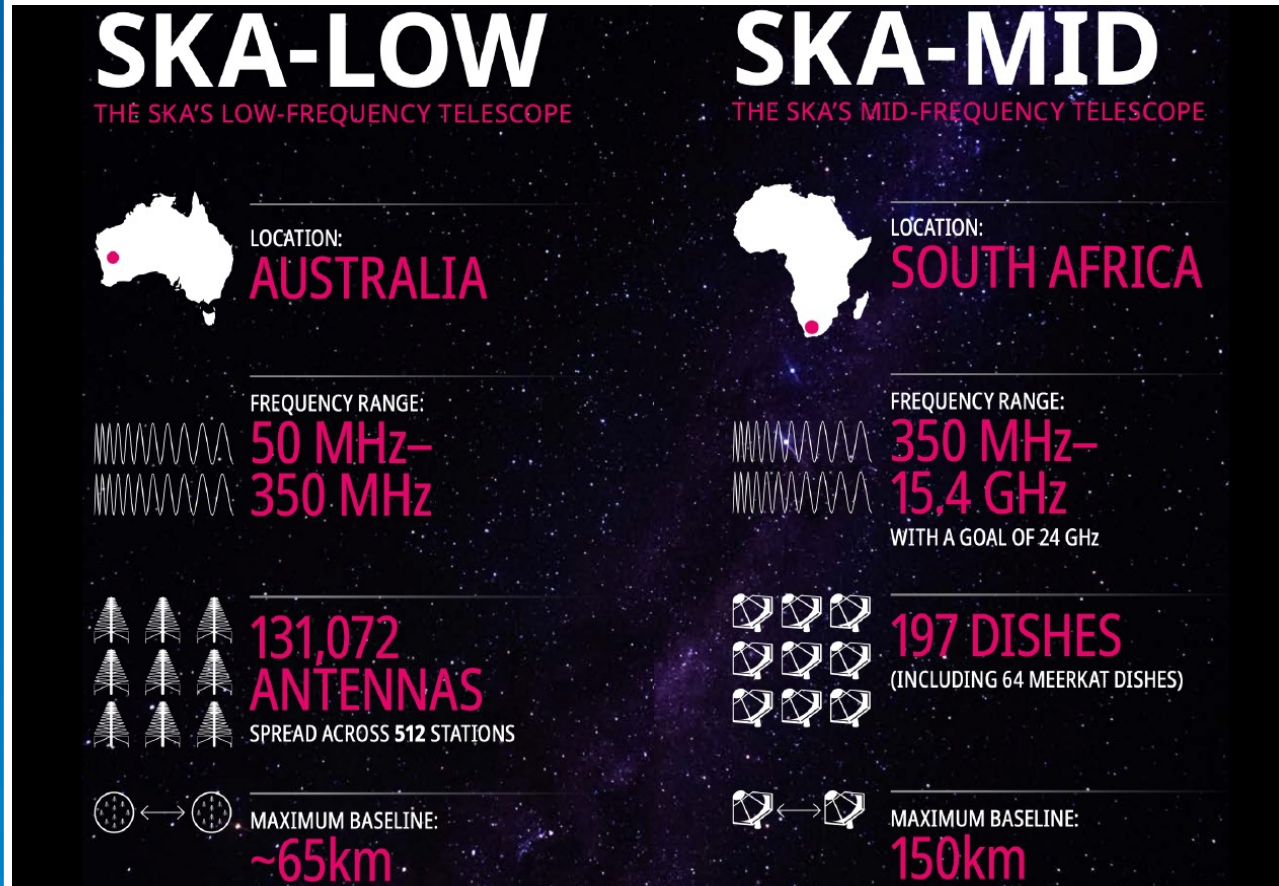
Image courtesy J. Swinbank

# Preparing for LOFAR2.0: Data Life Cycle & Early Cycle Data Retirement

Product type	Example	Retention period
Raw	unprocessed vis.	Not retained
Instrumental	Flagged, compressed vis.	18 months
Intermediate	Direction-independent vis	18 months
Advanced	Images, cubes	Indefinite
Special cases	Unique observations	For discussion

- LOFAR2 will generate considerably more data than LOFAR1: ~70 PB intermediate + ~30 PB advanced
- Data challenge outstrips current affordable solutions
- ILT-board approved a data life cycle:
  - Advanced data products (images, cubes, catalogues) kept indefinitely
  - Intermediate data products will be retired after a period (~18 months), based on available resources
  - Exceptions to be considered in exceptional cases
- To prepare for LOFAR2, a first step is taken now: retirement of early LOFAR Cycle data (Cycle 0 till Cycle 6)
  - Timeline: end-2024

# The SKA Challenge: Later This Decade



- **Science-ready data** generated by the SKA observatory: **600 PB/year** – new magnitude for astronomical context
- Providing this scientific data repository represents a **technical challenge for discovery, analysis and exploitation tasks**
- Should include **big data lakes and change of paradigm for data access**
- SKAO data products provided to a **network of SKA regional centers (SRC's)** responsible for archiving & data curation, making the data available to the users, enabling scientific discovery

# To Conclude

- LOFAR is an important technological pathfinder for next-generation data-intensive radio astronomy
- We face **fundamental challenges related to the data handling** and this will become even more acute for the LOFAR upgrades
- Demand for putting in place policies that adequately prepare the infrastructure for the future